

Finding statistics you can trust

Jeff Leek

@jtleek

www.jtleek.com

Sources of statistical software

```
#####  
# An R function to make a personalized map of people you follow and who follow you on twitter.  
# R functions Copyright (C) 2011 Jeff Leek (jtleek@gmail.com), and the Simply Statistics Blog  
# (http://simplystatistics.tumblr.com, http://twitter.com/simplystats)  
#  
# This program is free software: you can redistribute it and/or modify  
# it under the terms of the GNU General Public License as published by  
# the Free Software Foundation, either version 3 of the License, or  
# (at your option) any later version.  
#  
# This program is distributed in the hope that it will be useful,  
# but WITHOUT ANY WARRANTY; without even the implied warranty of  
# MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the  
# GNU General Public License for more details, see <http://www.gnu.org/licenses/>.  
#  
#  
# These functions depend on the packages: twitterR, maps, geosphere, and RColorBrewer. It will  
# attempt to install them if they are not installed when you source this function. Care  
# should be used when using this function since the twitterR API has rate limiting in place.  
# If you have a large number of followers, or run the function many times, you may be  
# rate limited.  
#  
#  
# How to use:  
#   # Source the function  
#   source("http://biostat.jhsph.edu/~jleek/code/twitterMap.R")  
#  
#   # Make your twittermap  
#   twitterMap("simplystats")  
#  
#   #If your location can't be found or latitude longitude can't be calculated  
#   #choose a bigger city near you. The list of cities used by twitterMap  
#   #can be found like so:  
#   data(world.cities)  
#   grep("Baltimore",world.cities[,1])  
#  
#   # Then make the map using that big city  
#   twitterMap("simplystats",userLocation="Baltimore")  
#
```



This repository Search

Pull requests Issues Gist

www.github.com

jtleek / sva-devel

Unwatch 6

Star 4

Fork 7

Description

Website

Short description of this repository

Website for this repository (optional)

Save or Cancel

26 commits

1 branch

0 releases

4 contributors

branch: master sva-devel / +

Commit made by the Bioconductor Git-SVN bridge.

bioc-sync authored 27 days ago latest commit 4e9c7a2731

R	Made the following changes: 1) added unit tests for ComBat to check C...	2 months ago
man	Made several modifications to ComBat to streamline the design matrix ...	5 months ago
src	Commit made by the Bioconductor Git-SVN bridge.	7 months ago
tests	Made the following changes: 1) added unit tests for ComBat to check C...	2 months ago
vignettes	Made several modifications to ComBat to streamline the design matrix ...	5 months ago
.gitignore	Initial commit	11 months ago
DESCRIPTION	Commit made by the Bioconductor Git-SVN bridge.	27 days ago
NAMESPACE	fixed documentation of sva.check	8 months ago

Code

Issues 0

Pull requests 0

Wiki

Pulse

Graphs

Settings

HTTPS clone URL

https://github.com/

You can clone with HTTPS, SSH, or Subversion.

Clone in Desktop

[A](#) [B](#) [C](#) [D](#) [E](#) [F](#) [G](#) [H](#) [I](#) [J](#) [K](#) [L](#) [M](#) [N](#) [O](#) [P](#) [Q](#) [R](#) [S](#) [T](#) [U](#) [V](#) [W](#) [X](#) [Y](#) [Z](#)[CRAN](#)[Mirrors](#)[What's new?](#)[Task Views](#)[Search](#)[About R](#)[R Homepage](#)[The R Journal](#)[Software](#)[R Sources](#)[R Binaries](#)[Packages](#)[Other](#)[Documentation](#)[Manuals](#)[FAQs](#)[Contributed](#)[A3](#)[abc](#)[ABCanalysis](#)[abc.data](#)[abcdeFBA](#)[ABCExtremes](#)[ABCoptim](#)[ABCp2](#)[abctools](#)[abd](#)[abf2](#)[abind](#)[abn](#)[abundant](#)[acc](#)[accelerometry](#)[AcceptanceSampling](#)[ACCLMA](#)[accrual](#)[accrued](#)[ACD](#)[acepack](#)[acid](#)[acm4r](#)[ACNE](#)[acnr](#)

A3: Accurate, Adaptable, and Accessible Error Metrics for Predictive Models

Tools for Approximate Bayesian Computation (ABC)

Computed ABC Analysis

Data Only: Tools for Approximate Bayesian Computation (ABC)

ABCDE_FBA: A-Biologist-Can-Do-Everything of Flux Balance Analysis with this package

ABC Extremes

Implementation of Artificial Bee Colony (ABC) Optimization

Approximate Bayesian Computational model for estimating P2

Tools for ABC analyses

The Analysis of Biological Data

Load Gap-Free Axon ABF2 Files

Combine Multidimensional Arrays

Data Modelling with Additive Bayesian Networks

Abundant regression and high-dimensional principal fitted components

Processes Accelerometer Data

Functions for Processing Minute-to-Minute Accelerometer Data

Creation and evaluation of Acceptance Sampling Plans

ACC & LMA Graph Plotting

Bayesian Accrual Prediction

Visualization tools for partially accruing data

Categorical data analysis with complete or missing responses

ace() and avas() for selecting regression transformations

Analysing Conditional Distributions of Income

Align-and-Count Method comparisons of RFLP data

Affymetrix SNP Probe-Summarization using Non-Negative Matrix Factorization

Annotated Copy-Number Regions

Modelling dependence with multivariate Archimax (or any user-defined continuous) copulas

[Home](#) » [Bioconductor 3.1](#) » [Software Packages](#) » [sva](#)

sva

available all platforms downloads top 5% posts 6 / 2 / 3 / 2
in Bioc 3.53 years build ok commits 1.17

Surrogate Variable Analysis

Bioconductor version: Release (3.1)

The `sva` package contains functions for removing batch effects and other unwanted variation in high-throughput experiment. Specifically, the `sva` package contains functions for the identifying and building surrogate variables for high-dimensional data sets. Surrogate variables are covariates constructed directly from high-dimensional data (like gene expression/RNA sequencing/methylation/brain imaging data) that can be used in subsequent analyses to adjust for unknown, unmodeled, or latent sources of noise. The `sva` package can be used to remove artifacts in three ways: (1) identifying and estimating surrogate variables for unknown sources of variation in high-throughput experiments (Leek and Storey 2007 PLoS Genetics, 2008 PNAS), (2) directly removing known batch effects using ComBat (Johnson et al. 2007 Biostatistics) and (3) removing batch effects with known control probes (Leek 2014 biorXiv). Removing batch effects and using surrogate variables in differential expression analysis have been shown to reduce dependence, stabilize error rate estimates, and improve reproducibility, see (Leek and Storey 2007 PLoS Genetics, 2008 PNAS or Leek et al. 2011 Nat. Reviews Genetics).

Author: Jeffrey T. Leek <jtleek at gmail.com>, W. Evan Johnson <wej at bu.edu>, Hilary S. Parker <hiparker at jhsph.edu>, Elana J. Fertig <ejfertig at jhmi.edu>, Andrew E. Jaffe <ajaffe at jhsph.edu>, John D. Storey <jstorey at princeton.edu>

Maintainer: Jeffrey T. Leek <jtleek at gmail.com>, John D. Storey <jstorey at princeton.edu>, W. Evan Johnson <wej at bu.edu>

Workflows »

Common Bioconductor workflows include:

- [Oligonucleotide Arrays](#)
- [High-throughput Sequencing](#)
- [Counting Reads for Differential Expression](#) (parathyroideSE vignette)
- [Annotation](#)
- [Annotating Variants](#)
- [Annotating Ranges](#)
- [Flow Cytometry](#) and other assays
- [Candidate Binding Sites for Known Transcription Factors](#)
- [Cloud-enabled cis-eQTL search and annotation](#)
- [RNA-Seq workflow: gene-level exploratory analysis and differential expression](#)
- [Changing genomic coordinate systems with rtracklayer::liftOver](#)
- [Mass spectrometry and proteomics data analysis](#)

Mailing Lists »

Post questions about Bioconductor packages to our mailing lists. Read the [posting guide](#) before posting!

- [bioconductor](#)
- [bioc-devel](#)

Measuring trust

Home » [Bioconductor 3.1](#) » [Software Packages](#) » [sva](#)

sva

available all platforms downloads top 5% posts 6 / 2 / 3 / 2
in Bioc 3.53 years build ok commits 1.17

Surrogate Variable Analysis

Bioconductor version: Release (3.1)

The `sva` package contains functions for removing batch effects and other unwanted variation in high-throughput experiment. Specifically, the `sva` package contains functions for the identifying and building surrogate variables for high-dimensional data sets. Surrogate variables are covariates constructed directly from high-dimensional data (like gene expression/RNA sequencing/methylation/brain imaging data) that can be used in subsequent analyses to adjust for unknown, unmodeled, or latent sources of noise. The `sva` package can be used to remove artifacts in three ways: (1) identifying and estimating surrogate variables for unknown sources of variation in high-throughput experiments (Leek and Storey 2007 PLoS Genetics, 2008 PNAS), (2) directly removing known batch effects using ComBat (Johnson et al. 2007 Biostatistics) and (3) removing batch effects with known control probes (Leek 2014 biorXiv). Removing batch effects and using surrogate variables in differential expression analysis have been shown to reduce dependence, stabilize error rate estimates, and improve reproducibility, see (Leek and Storey 2007 PLoS Genetics, 2008 PNAS or Leek et al. 2011 Nat. Reviews Genetics).

Author: Jeffrey T. Leek <jtleek at gmail.com>, W. Evan Johnson <wej at bu.edu>, Hilary S. Parker <hiparker at jhsph.edu>, Elana J. Fertig <ejfertig at jhmi.edu>, Andrew E. Jaffe <ajaffe at jhsph.edu>, John D. Storey <jstorey at princeton.edu>

Maintainer: Jeffrey T. Leek <jtleek at gmail.com>, John D. Storey <jstorey at princeton.edu>, W. Evan Johnson <wej at bu.edu>

Downloads

- [High-throughput Sequencing](#)
- [Counting Reads for Differential Expression](#) (parathyroideSE vignette)
- [Annotation](#)
- [Annotating Variants](#)
- [Annotating Ranges](#)
- [Flow Cytometry](#) and other assays
- [Candidate Binding Sites for Known Transcription Factors](#)
- [Cloud-enabled cis-eQTL search and annotation](#)
- [RNA-Seq workflow: gene-level exploratory analysis and differential expression](#)
- [Changing genomic coordinate systems with rtracklayer::liftOver](#)
- [Mass spectrometry and proteomics data analysis](#)

Mailing Lists »

Post questions about Bioconductor packages to our mailing lists. Read the [posting guide](#) before posting!

- [bioconductor](#)
- [bioc-devel](#)

Home » [Bioconductor 3.1](#) » [Software Packages](#) » [sva](#)

sva

available all platforms downloads top 5% posts 6 / 2 / 3 / 2
in Bioc 3.53 years build ok commits 1.17

Responsiveness

Surrogate Variable Analysis

Bioconductor version: Release (3.1)

The `sva` package contains functions for removing batch effects and other unwanted variation in high-throughput experiment. Specifically, the `sva` package contains functions for the identifying and building surrogate variables for high-dimensional data sets. Surrogate variables are covariates constructed directly from high-dimensional data (like gene expression/RNA sequencing/methylation/brain imaging data) that can be used in subsequent analyses to adjust for unknown, unmodeled, or latent sources of noise. The `sva` package can be used to remove artifacts in three ways: (1) identifying and estimating surrogate variables for unknown sources of variation in high-throughput experiments (Leek and Storey 2007 PLoS Genetics, 2008 PNAS), (2) directly removing known batch effects using ComBat (Johnson et al. 2007 Biostatistics) and (3) removing batch effects with known control probes (Leek 2014 biorXiv). Removing batch effects and using surrogate variables in differential expression analysis have been shown to reduce dependence, stabilize error rate estimates, and improve reproducibility, see (Leek and Storey 2007 PLoS Genetics, 2008 PNAS or Leek et al. 2011 Nat. Reviews Genetics).

Author: Jeffrey T. Leek <jtleek at gmail.com>, W. Evan Johnson <wej at bu.edu>, Hilary S. Parker <hiparker at jhsph.edu>, Elana J. Fertig <ejfertig at jhmi.edu>, Andrew E. Jaffe <ajaffe at jhsph.edu>, John D. Storey <jstorey at princeton.edu>

Maintainer: Jeffrey T. Leek <jtleek at gmail.com>, John D. Storey <jstorey at princeton.edu>, W. Evan Johnson <wej at bu.edu>

- [High-throughput Sequencing](#)
- [Counting Reads for Differential Expression](#) (parathyroideSE vignette)
- [Annotation](#)
- [Annotating Variants](#)
- [Annotating Ranges](#)
- [Flow Cytometry](#) and other assays
- [Candidate Binding Sites for Known Transcription Factors](#)
- [Cloud-enabled cis-eQTL search and annotation](#)
- [RNA-Seq workflow: gene-level exploratory analysis and differential expression](#)
- [Changing genomic coordinate systems with rtracklayer::liftOver](#)
- [Mass spectrometry and proteomics data analysis](#)

Mailing Lists »

Post questions about Bioconductor packages to our mailing lists. Read the [posting guide](#) before posting!

- [bioconductor](#)
- [bioc-devel](#)



This repository Search

Pull requests Issues Gist

www.github.com

jtleek / sva-devel

Unwatch 6

★ Star 4

Fork 7

Description

Short description of this repository

Other people like it

26 commits

1 branch

0 releases

4 contributors

branch: master sva-devel / +

Commit made by the Bioconductor Git-SVN bridge.

bioc-sync authored 27 days ago	latest commit 4e9c7a2731
R	Made the following changes: 1) added unit tests for ComBat to check C... 2 months ago
man	Made several modifications to ComBat to streamline the design matrix ... 5 months ago
src	Commit made by the Bioconductor Git-SVN bridge. 7 months ago
tests	Made the following changes: 1) added unit tests for ComBat to check C... 2 months ago
vignettes	Made several modifications to ComBat to streamline the design matrix ... 5 months ago
.gitignore	Initial commit 11 months ago
DESCRIPTION	Commit made by the Bioconductor Git-SVN bridge. 27 days ago
NAMESPACE	fixed documentation of sva.check 8 months ago

Code

Issues 0

Pull requests 0

Wiki

Pulse

Graphs

Settings

HTTPS clone URL

https://github.com/

You can clone with [HTTPS](#), [SSH](#), or [Subversion](#).

Clone in Desktop



This repository Search

Pull requests Issues Gist

www.github.com

jtleek / sva-devel

Unwatch 6

Star 4

Fork 7

Description

Short description of this repository

26 commits

People have been working on it

branch: master sva-devel / +

Commit made by the Bioconductor Git-SVN bridge.

bioc-sync	authored 27 days ago	latest commit 4e9c7a2731
R	Made the following changes: 1) added unit tests for ComBat to check C...	2 months ago
man	Made several modifications to ComBat to streamline the design matrix ...	5 months ago
src	Commit made by the Bioconductor Git-SVN bridge.	7 months ago
tests	Made the following changes: 1) added unit tests for ComBat to check C...	2 months ago
vignettes	Made several modifications to ComBat to streamline the design matrix ...	5 months ago
.gitignore	Initial commit	11 months ago
DESCRIPTION	Commit made by the Bioconductor Git-SVN bridge.	27 days ago
NAMESPACE	fixed documentation of sva.check	8 months ago

Code

Issues 0

Pull requests 0

Wiki

Pulse

Graphs

Settings

HTTPS clone URL

https://github.com/

You can clone with HTTPS, SSH, or Subversion.

Clone in Desktop

Where you can ask about methods



Community

User Login

New Post

or Classic search

Limit to: all time

<prev • 24,531 results • page 1 of 614 • next >

Sort by: update

- 0** votes **1** answer **43** views **Problem with Circos ideogram radius below 1**
circos written 53 minutes ago by michalzwielokopolski • 0 • updated 1 minute ago by John • 710
- 0** votes **0** answers **16** views **Difference of kinds of FPKM between cufflink and cuffdiff**
rna-seq written 17 minutes ago by tekitekito • 0
- 3** votes **1** answer **138** views **Spliced aligners for RNA-seq**
alignment next-gen rna-seq written 16 hours ago by amirmhzadeh • 30 • updated 26 minutes ago by h.mon • 200
- 0** votes **1** answer **85** views **error with trinity installation**
error with trinity installation written 5 hours ago by Kurban • 80 • updated 41 minutes ago by h.mon • 200
- 0** votes **1** answer **39** views **Hey any one used QC-Chain tool?**
qc-chain written 2 hours ago by gskbioinfo143 • 10 • updated 54 minutes ago by h.mon • 200
- 0** votes **2** answers **58** views **Comparison of sequence homology of two genes**
sequence homology webtool written 1 hour ago by Jimbou • 280 • updated 1 hour ago by garimaverma311 • 0
- 0** votes **0** answers **40** views **How to read the GenBank (.gbk) file and output the following extracted features in tabular format: Start and Stop position, Strand, protein_id, Locus_tag and product**
gene genome sequence written 1 hour ago by Sumit • 0
- 0** votes **0** answers **47** views **sequence in forward seq files but not in the reverse files**
next-gen snp sequencing written 1 hour ago by flavobacteria • 20
- 12** votes **2** answers **126** views **Bioinformatics online course**
course teaching written 3 hours ago by ashel123 • 40 • updated 1 hour ago by shelly.deforte • 120
11 follow
- 0** votes **1** answer **124** views **quality control of RNA-seq results**
tool rna-seq written 15 hours ago by elmira.mohyedin • 10 • updated 2 hours ago by Devon Ryan • 32k

Recent Votes

- C: The new era of bioinformatics: simple and fast tools, fewer and more informative
- Bioinformatics online course
- C: Running and Analyzing fastqc on multiple fastq files
- C: Running and Analyzing fastqc on multiple fastq files
- C: Spliced aligners for RNA-seq
- A: ensembl biomart: why some genes are not found?
- A: Bioinformatics online course

Recent Locations • All »

- Germany, 1 minute ago
- New York, NY, United States, 1 minute ago
- UK, 3 minutes ago
- Serbia And Montenegro, 3 minutes ago
- Italy, 3 minutes ago
- United Kingdom, 3 minutes ago
- United States, 4 minutes ago

Recent Awards • All »

- Scholar 🏆 to Sean Davis • 18k
- Appreciated ❤️ to venu • 380
- Good Answer 🏆 to venu • 380
- Scholar 🏆 to venu • 380
- Student 🎓 to ashel123 • 40
- Teacher 🏆 to venu • 380

Recent Replies

- C: More about nanobody? by h.mon • 200

I think this question is more appropriate



[ASK QUESTION](#)

[LATEST](#)

[NEWS](#)

[JOBS](#)

[TAGS](#)

[USERS](#)

Limit ▾ Sort ▾ Search

0 votes **0** answers **7** views [txdb](#) [genomicfeatures](#) **Setting package name in makeTxDbPackage** written 3 hours ago by [Diego Diez](#) • 480

0 votes **0** answers **7** views [txdb](#) [genomicfeatures](#) **Setting genome information on a TxDb object** written 3 hours ago by [Diego Diez](#) • 480

0 votes **0** answers **10** views [affymetrix](#) [eisa](#) [isa2](#) [microarray](#) **Remove coexpressed gene pair modules using eisa package** written 8 hours ago by [dhwani.dholakia](#) • 0

0 votes **0** answers **14** views [diffbind](#) **Diffbind.report error too many NAs** written 13 hours ago by [mm2489](#) • 0

2 votes **2** answers **54** views [formula](#) [maseq](#) [deseq2](#) [limma](#) **Design formula and Design matrix in DESeq** written 22 hours ago by [aurelie.bornot](#) • 10 • updated 13 hours ago by [Gordon Smyth](#) • 22k

3 votes **3** answers **199** views [limma](#) [microarray](#) **Annotating the results with associated gene symbols** written 6 months ago by [sss1991.f](#) • 0 • updated 14 hours ago by [Gordon Smyth](#) • 22k

1 vote **3** answers **219** views [limma](#) **annotating the results with association gene symbol** written 13 months ago by [Guest User](#) • 12k • updated 14 hours ago by [Gordon Smyth](#) • 22k

0 votes **1** answer **31** views [differential expression](#) [differential gene expression](#) [edger](#) **Find differentially expressed values** written 20 hours ago by [Safana](#) • 0 • updated 15 hours ago by [Gordon Smyth](#) • 22k

0 votes **3** answers **29** views [maseq](#) [normalization](#) [tcga](#) [deseq2](#) **Different clustering results of DESeq2 counts and rlog** written 19 hours ago by [ezz](#) • 0

0 votes **0** answers **12** views [crlmm](#) [vanillaice crlmm](#) [cnvtools](#) [vanillaice](#) [cnv](#) **Why would CRLMM produces all NAs for log R ratios and B Allele frequencies?** written 19 hours ago by [cafelumiere12](#) • 0

1 vote **1** answer **28** views [log](#) [deseq2](#) **Dependence of rlog transformed value range on number of samples** written 20 hours ago by [michael](#) • 1.5k

Recent...

Replies

- [C: Design formula and Desig...](#) by [aurelie.bornot](#) • 10
- [C: Design formula and Desig...](#) by [aurelie.bornot](#) • 10
- [C: DESeq2 with 2 factors: p...](#) by [aurelie.bornot](#) • 10
- [C: Can I use the WGCNA dend...](#) by [Jon Bråte](#) • 120
- [C: Can I use the WGCNA dend...](#) by [Dario Strbenac](#) • 1.1k

Votes

- [A: DESeq2 comparison with m...](#)
- [A: Design formula and Desig...](#)
- [A: Design formula and Desig...](#)
- [annotating the results with...](#)
- [A: DESeq2 results with RUVS...](#)

Awards • All »

- [Teacher](#) 🏆 to [Michael Love](#) • 4.5k
- [Scholar](#) 🏆 to [Michael Love](#) • 4.5k
- [Scholar](#) 🏆 to [Michael Love](#) • 4.5k
- [Commentator](#) 🗨️ to [James W. MacDonald](#) • 35k
- [Scholar](#) 🏆 to [andrew.j.skilton73](#) • 70
- [Scholar](#) 🏆 to [Michael Love](#) • 4.5k

Locations • All »

- [United States](#), 6 minutes ago
- [United States](#), 13 minutes ago
- [United States](#), 28 minutes ago
- [Japan](#), 38 minutes ago
- [United States](#), 42 minutes ago